

The Exponentiated Complementary Exponential Geometric Distribution

ECEG Distribution

Cíntia Y. Yamachi¹, Mari Roman², Francisco Louzada³, Maria A. P. Franco⁴, Vicente G. Cancho⁵

^{1,2,4} Departamento de Estatística, Universidade Federal de São Carlos/DEs, ^{3,5} Instituto de Ciências Matemáticas e Computação, Universidade de São Paulo/ICMC
São Carlos, SP - Brasil

¹cintia.yurie@yahoo.com.br; ²mari.roman19@gmail.com; ³louzada@icmc.usp.br; ⁴mapfranco@ufscar.br;

⁵garibay@icmc.usp.br

Abstract

Recently new classes of models have appeared to model survival times, such as, Exponentiated Exponential distribution (Gupta & Kundu, 1999), an extension of Lidley distribution (Bakouch, Al-Zahrani, Al-Shomrani, Marchi, and Louzada, 2011), a generalization of Frechét, gamma, Gumbel (Nadarajah & Kotz, 2006), which can accommodate increasing, decreasing, unimodal hazard functions. In this paper it discussed a generalization of a distribution presented by Louzada, Roman and Cancho (2011), which arise on latent competing risks scenarios, only the maximum lifetime is observed among all risks, that can accommodates increasing hazard functions. The density, survival function and failure rate are presented. Statistical proprieties as characteristic function, moments, r -th order moment, quantile function, residual lifetime distribution are provided. The maximum likelihood estimator and inference from the distribution are discussed. The performance in a real dataset is verified by modeling different distributions to the survival times.

Keywords

Complementary Risks; Decreasing Failure Rate; Increasing Failure Rate; Exponentiated Distributions; Maximum Likelihood Estimation

Introduction

The Complementary Exponential Geometric (CEG) distribution was introduced by Louzada, Roman and Cancho (2011), this distribution presents on a latent complementary risks scenarios, in a situation with several factor which the cause of the failure is unknown, only the maximum lifetime value among all risks is observed, characterizing a complementary risks (CR) problem (Basu and Klein, 1982). In the classical CR scenarios the lifetime associated with a particular risk is not observable, rather we observe only the maximum lifetime value among all risks.

Simplistically, in reliability, we observe only the maximum component lifetime of a parallel system. That is, the observable quantities for each component are the maximum lifetime values to failure among all risks, and the cause of failure. The CR dual are the so called competing risks scenarios, where the lifetime associated with a particular risk is not observable, rather we observe only the minimum lifetime values among all risks. Full statistical procedures and extensive literature are available to deal with these problems and interested readers can refer to Lawless (2003), Crowder, Kimber, Smith and Sweeting (1991) and Cox & Oakes (1984).

In scenario, compositions of distributions, there are several works, how for instance, Adamidis & Loukas (1998) with Exponential Geometric distribution, the Poisson Exponential distribution introduced by Cancho, Louzada and Barriga (2011), Silva, Ortega and Cordeiro (2010) studied in details the beta modified Weibull distribution and Silva, Barreto-Souza and Cordeiro (2010) defined the Generalized Exponential Geometric distribution and the Weibull Geometric distribution was proposed by Barreto-Souza, Morais and Cordeiro (2010).

Generalization of the the Exponential distribution is a widely used lifetime distribution for modeling many problems in lifetime testing and reliability studies. In recent years, several new classes of models were introduced grounded in its simple, elegant and close form, such as Gupta & Kundu (1999), Barreto-Souza & Cribari-Neto (2009), and also the precursor of the CEG, the Exponential Geometric distribution presented by Adamidis & Loukas (1998).

The exponentiation of distributions is a mechanism that makes the model more flexible, Nadarajah & Kotz (2006) introduce four more exponentiated type distributions: the Exponentiated Gamma,

Exponentiated Weibull, exponentiated Gumbel and the Exponentiated Fréchet distribution. We also, several authors presented exponentiated distributions, such as Mudholkar, Srivastava and Freimer (1995) with the exponentiation of the Weibull distribution, Gupta & Kundu (1999) with the Exponentiated Exponential, Barriga, Louzada and Cancho (2011) with the Complementary Exponential Power distribution which is the exponentiation of the Exponential Power distribution proposed by Smith & Bain (1975) denoted as Complementary Exponential Power distribution, Bakouch, Al-Zahrani, Al-Shomrani, Marchi and Louzada (2011) with the extension of the Lindley (EL) distribution and the Complementary Exponential Power distribution (CEP) introduced by Barriga, Louzada and Cancho (2011).

In this paper, we propose a new lifetime distribution family, which is a generalization of the CEG distribution, namely the family Exponentiated Complementary Exponential Geometric (ECEG) distribution. This distribution is obtained by exponentiating the cumulate density function of CEG. It has two parameters λ and θ of scale and shape respectively, its density function and cumulative density are given in by

$$f_{baseline}(y) = \frac{\lambda \theta e^{-\lambda y}}{(e^{-\lambda y}(1-\theta) + \theta)^2}, \quad (1)$$

$$F_{baseline}(y) = \frac{\theta(1 - e^{-\lambda y})}{e^{-\lambda y}(1-\theta) + \theta}, \quad (2)$$

where, $y > 0$ is the time, $0 < \theta < 1$ and $\lambda > 0$.

The CEG distribution accommodates increasing failure while a new family ECEG one accommodates also decreasing failure rate besides increasing ones. The paper is organized as following. Section 2 presents the density, survival, cumulate and hazard function for this family. The proprieties as characteristic function, moments, r -th order moment, quantile function, residual lifetime distribution are presented in Section 3. Section 4 and Section 5 present respectively the inference method and application in artificial and real dataset. Section 6 provides some concluding remarks.

The Eceg Distribution

Marshall & Olkin (2007) presented the idea of exponentiation, based in the baseline cumulative distribution function $F_{baseline}(y)$ and an arbitrary power $\alpha > 0$, obtaining a new cumulate distribution function $F(y) = [F_{baseline}(y)]^\alpha$, where α can be referred as a resilience parameter and $F(y)$ is a resilience parameter family. The Exponentiated Complementary

Exponential Geometric (ECEG) is obtain exponentiating the cumulate density function from CEG distribution (2) proposed by Louzada, Roman, and Cancho (2011), and it is given by,

$$F(y) = \left(\frac{\theta(1 - e^{-\lambda y})}{e^{-\lambda y}(1-\theta) + \theta} \right)^\alpha, \quad (3)$$

where $y > 0$, $\alpha > 0$, $0 < \theta < 1$ and $\lambda > 0$.

The density function of a random variable with ECEG distribution is given by:

$$f(y) = \alpha \left(\frac{\theta(1 - e^{-\lambda y})}{e^{-\lambda y}(1-\theta) + \theta} \right)^{\alpha-1} \left(\frac{\theta \lambda e^{-\lambda y}}{(e^{-\lambda y}(1-\theta) + \theta)^2} \right). \quad (4)$$

In (4) α and θ are shape parameters and λ is the scale parameter. For $\alpha = 1$ the ECEG is reduced to the CEG distribution Louzada, Roman, and Cancho (2011). For $\alpha = \theta = 1$ the ECEG reduced to e Exponential (E) distribution.

The survival function of a ECEG distribution random variable is given by

$$S(y) = 1 - \left(\frac{\theta(1 - e^{-\lambda y})}{e^{-\lambda y}(1-\theta) + \theta} \right)^\alpha. \quad (5)$$

And from (4) and (5) the hazard function, according to the relationship $h(y) = f(y)/S(y)$, is given by

$$h(y) = \frac{\alpha \lambda e^{-\lambda y}}{(e^{-\lambda y}(1-\theta) + \theta)(1 - e^{-\lambda y})} \frac{\theta(1 - e^{-\lambda y})^\alpha}{(e^{-\lambda y}(1-\theta) + \theta)^\alpha - (\theta(1 - e^{-\lambda y}))^\alpha}. \quad (6)$$

And it can be increasing and decreasing as shown in the Figure 1.

The ECEG distribution accommodates, unimodal, and a broad variety of monotone hazard functions depend on the parameter values over the regions of the space of the shape parameters.

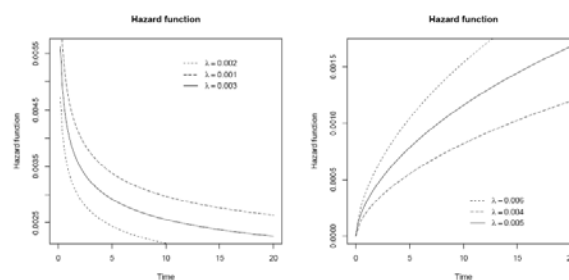


Figure 1 - Hazard function of the ECEG distribution. Left panel: for $\alpha = 0.90$ and $\theta = 0.50$. Right panel: for $\alpha = 1.60$ and $\theta = 0.95$

(i) If $\alpha > 1$, $\alpha\theta > 1$ or $\alpha\theta < 1$, and $\lambda\theta < 1$ we have $h(y)$ increasing.

(ii) If $\alpha < 1$, $\alpha\theta < 1$, and $\lambda\theta < 1$ we have $h(y)$ decreasing.

Since the hazard rate function is complex, the shape of the hazard function was obtained by composition of conditions, to compare with the hypotheses from Glaser's theorem (Glaser, 1980), which talks about this subject by the following considerations: Define $\eta(y) = f'(y)/f(y)$, where $f'(y)$ is the first derivative of the (4).

Hence,

$$\eta(y) = -(\alpha - 1) \frac{\lambda e^{-\lambda y} - 2e^{-\lambda y} - \theta + 1 + \theta e^{-\lambda y}}{(1 - e^{-\lambda y}) e^{-\lambda y} (1 - \theta) + \theta},$$

$$\eta'(y) = -(\alpha - 1) \left[\frac{-\lambda^2 e^{-\lambda y} (1 - e^{-\lambda y}) + \lambda^2 e^{-2\lambda y}}{(1 - e^{-\lambda y})^2} \right]$$

$$\times \left(\frac{e^{-\lambda y} (-2 + \theta) + 1 - \theta}{e^{-\lambda y} (1 - \theta) + \theta} \right) - (\alpha - 1) \frac{\lambda e^{-\lambda y}}{1 - e^{-\lambda y}}$$

$$\times \left[\frac{-\lambda e^{-\lambda y} (\theta - 2)(e^{-\lambda y} (1 - \theta) + \theta) + \lambda (1 - \theta) e^{-\lambda y} (e^{-\lambda y} (-2 + \theta) + 1 - \theta)}{(e^{-\lambda y} (1 - \theta) + \theta)^2} \right].$$

And, this result, we have the conditions listed above, where if $\alpha > 1$, $\alpha\theta > 1$ or $\alpha\theta < 1$, and $\lambda\theta < 1$, for every $y > 0$, $\eta'(y) > 0$, consequently the hazard function $h(y)$ is increasing and if $\alpha < 1$, $\alpha\theta < 1$, and $\lambda\theta < 1$ for every $y > 0$, $\eta'(y) < 0$, therefore the failure rate function is decreasing.

Some Properties

Many of the most important features and characteristics of a distribution can be studied through its moments, such mean, variance. A general expression for r -th $\mu_r = E(Y^r)$ ordinary moment of the ECEG is hard to be obtained and we resume the mean and variance, as it follows. Moment-generating of the Y variable, with density function given by (3) can be obtained analytically.

For any real number t , let $\Phi_Y(t)$ be the characteristic function of Y , that is, $\Phi_Y(t) = E(e^{ity})$, where i denotes the imaginary unit. With the preceding notations, we state the following.

Proposition 1: For the random variable Y with ECEG distribution, we have that, its characteristic function is given by

$$\Phi_Y(t) = \alpha \theta^\alpha \Gamma \left(\frac{-it}{\lambda} + 1 \right) \sum_{k=0}^{\infty} \binom{-(\alpha+1)}{k} \times \frac{(\theta-1)^k \Gamma(\alpha+k)}{\Gamma(\alpha+1 - \frac{it}{\lambda} + k)}, \quad (7)$$

where $i = \sqrt{-1}$.

Proof 1:

$$\Phi_Y(t) = \int_0^{\infty} e^{ity} f(y) dy$$

$$= \int_0^{\infty} e^{y(it-\lambda)} \alpha \theta \lambda \left(\frac{\theta(1-e^{-\lambda y})}{e^{-\lambda y}(1-\theta) + \theta} \right)^\alpha \frac{e^{-\lambda y}}{\theta(1-e^{-\lambda y})} \frac{1}{(e^{-\lambda y}(1-\theta) + \theta)} dy$$

$$= \int_0^1 (1-z)^{\frac{-it}{\lambda}+1} \alpha \theta^\alpha z^{\alpha-1} \frac{1}{(1-(1-\theta)z)^{\alpha+1}} dz,$$

where the last equality follows from the change of variable $z = 1 - e^{-\lambda y}$.

Comparing the last integral with (17), obtain: $n = -it/\lambda + 1$, $b = (\theta - 1)$, $m = 1$, $p = \alpha$, $l = -(\alpha + 1)$, and making the appropriate substitutions, completed the proof.

Proposition 2: If Y has ECEG distribution, then

$$E(Y) = -\frac{\alpha \theta^\alpha \Gamma(\alpha)}{\lambda \Gamma(\alpha+1)} \frac{{}_2F_1^{(0,0,1,0)}(\alpha, \alpha+1, \alpha+1, 1-\theta) - \theta^\alpha H_\alpha}{\lambda \Gamma(\alpha+1)}, \quad (8)$$

and

$$V(Y) = \frac{\alpha \theta^\alpha \Gamma(\alpha) (-2 {}_2F_1^{(0,0,1,0)}(v) H_\alpha + {}_2F_1^{(0,0,2,0)}(v))}{\lambda^2 \Gamma(\alpha+1)}$$

$$+ \frac{\theta^{-\alpha} \left(\frac{\pi^2}{6} + H(\alpha)^2 - \Psi^{(1)}(1, 1 + \alpha) \right)}{\lambda^2 \Gamma(\alpha+1)} - \frac{\alpha^2 \theta^{2\alpha} \Gamma(\alpha)^2 ({}_2F_1^{(0,0,1,0)}(v) - \theta^{-\alpha} H_\alpha)^2}{\lambda^2 \Gamma(\alpha+1)^2}, \quad (9)$$

where,

$$H_\alpha = \sum_{k=1}^{\alpha} \frac{1}{k}, \quad \Psi^{(n)}(z) = \frac{d^{n+1}}{dz^{n+1}} \ln[\Gamma(z)], \quad v = (\alpha, \alpha+1, \alpha+1, 1-\theta)$$

and ${}_2F_1^{(0,0,1,0)}(v)$ denotes the hypergeometric function, given by

$${}_2F_1(a, b, c, z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!}$$

Proof 2: Setting $t=0$ in (7) and considering that $E(Y^r) =$

$\Phi_Y(0)^{(r)}/i^r$, we have the first and second moments. By the first and second moment the variance can be easily found.

The p -th quantile of the extended Complementary Exponential Geometric, the inverse of $F(y_p) = p$, is given by

$$Q(p) = F^{-1}(p) = -\lambda^{-1} \ln \left(\frac{1 - p^{\frac{1}{\alpha}}}{\frac{p^{\frac{1}{\alpha}}}{\theta} - p^{\frac{1}{\alpha}} + 1} \right), \quad (10)$$

where p has the Uniform(0,1) distribution.

Let Y_1, Y_2, \dots, Y_n a random sample take from ECEG distribution. Denote $Y_{1:n}, \dots, Y_{n:n}$ the order statistics.

Following the density function of i -th order statistic:

$$f_{i:n}(y) = \frac{n!}{(k-1)!(n-k)!} \left[\left(\frac{\theta(1-e^{-\lambda y})}{e^{-\lambda y}(1-\theta) + \theta} \right)^\alpha \right]^{k-1}$$

$$\times \left[1 - \left(\frac{\theta(1-e^{-\lambda y})}{e^{-\lambda y}(1-\theta) + \theta} \right)^\alpha \right]^{n-k} \alpha \left(\frac{1}{\theta(1-e^{-\lambda y})} \right) \left(\frac{\theta \lambda e^{-\lambda y}}{e^{-\lambda y}(1-\theta) + \theta} \right).$$

The r -th moment of the i -th order statistic $Y_{i:n}$ can be obtained from the known result,

$$E[Y_{i:n}^r] = r \sum_{p=n-i+1}^n (-1)^{p-n+i-1} \binom{p-1}{n-i} \binom{n}{p} \int_0^{\infty} y^{r-1} [S(y)]^p dy. \quad (11)$$

Proposition 3: For the random variable Y with ECEG distribution, we have that, r -th moment of the i -th order statistic is given by

$$E[Y_{i:n}^r] = \frac{r!}{\lambda^{r-1}} \sum_{p=n-i+1}^n \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} (-1)^{2(r-1)+p-n+i-1} \binom{p-1}{n-i} \binom{n}{p}$$

$$\times \frac{(-p)_k (-1)_l \theta^{\alpha k} (1-\theta)^m (\alpha k)_m}{k! l! m! n! (n+1)^r} [-(\alpha k + m + l)]_n. \quad (12)$$

Proof 3:

$$\begin{aligned}
\int_0^{\infty} y^{r-1} [S(y)]^p dy &= \int_0^{\infty} y^{r-1} \left[1 - \left(\frac{\theta(1-e^{-\lambda y})}{e^{-\lambda y}(1-\theta) + \theta} \right)^{\alpha} \right]^p dy \\
&= (-1)^{r-1} \left(\frac{1}{\lambda} \right)^{r-1} \sum_{k=0}^{\infty} \frac{(-p)_k}{k!} \sum_{l=0}^{\infty} \frac{(-1)_l}{l!} \theta^{\alpha k} (1-\theta)^m \\
&\times \sum_{m=0}^{\infty} \frac{(\alpha k)_m}{m!} \int_0^1 u^{\alpha k+m+l} (\ln(1-u))^{r-1} du = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} (-1)^{2(r-1)} \left(\frac{1}{\lambda} \right)^{r-1} \\
&\times \frac{(-p)_r (-1)_l \theta^{\alpha k} (1-\theta)^m (\alpha k)_m}{k! l! m! n! (n+1)^r} (-[\alpha k + m + l]_n (r-1)!).
\end{aligned}$$

From (11) and (18), the result follows.

The residual lifetime distribution of a random variable Y , distributed as ECEG, has the survival function:

$$S_r(y) = \Pr[Y > y + t | Y > t] = \frac{1 - \left(\frac{\theta(1-e^{-\lambda(y+t)})}{e^{-\lambda(y+t)}(1-\theta) + \theta} \right)^{\alpha}}{1 - \left(\frac{\theta(1-e^{-\lambda y})}{e^{-\lambda y}(1-\theta) + \theta} \right)^{\alpha}}. \quad (13)$$

The mean residual lifetime of a continuous distribution with survival function $S(t)$ is given by:

$$\mu(t) = E(Y - t | Y > y) = \frac{1}{S(t)} \int_t^{\infty} S(u) du. \quad (14)$$

Proposition 4: For the random variable Y with ECEG distribution, we have that, the mean residual lifetime is given by

$$\begin{aligned}
\mu(t) &= \frac{1}{1 - \left(\frac{\theta(1-e^{-\lambda t})}{e^{-\lambda t}(1-\theta) + \theta} \right)^{\alpha}} \left(\frac{1}{\lambda \theta^{\alpha}} \sum_{k=0}^{\infty} \frac{(e^{-\lambda t}(1-\theta) + \theta)^{k+1} - \theta^{k+1}}{(k+1)!} \right. \\
&+ \left. \left(\frac{\theta}{\theta-1} \right)^{\alpha} \frac{1}{\lambda} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \frac{(-1)_l ((e^{-\lambda t}(1-\theta) + \theta)^{l+m-\alpha+1} - \theta^{l+m-\alpha+1})}{l! m! \theta^m (l+m-\alpha+1)} \right). \quad (15)
\end{aligned}$$

Proof 4: From (14) and using $S(y)$ given by (5) we have that

$$\begin{aligned}
\mu(t) &= \frac{1}{S(t)} \int_t^{\infty} \left[1 - \left(\frac{\theta(1-e^{-\lambda u})}{e^{-\lambda u}(1-\theta) + \theta} \right)^{\alpha} \right] du \\
&= \frac{1}{S(t)} \left(\int_0^{e^{-\lambda t}(1-\theta) + \theta} \left[1 - \left(\frac{\theta \left(1 - \frac{a-\theta}{1-\theta} \right)}{a} \right)^{\alpha} \right] \times \frac{1}{(\theta-a)\lambda} da \right)
\end{aligned}$$

$$\begin{aligned}
\mu(t) &= \frac{1}{1 - \left(\frac{\theta(1-e^{-\lambda t})}{e^{-\lambda t}(1-\theta) + \theta} \right)^{\alpha}} \left(\frac{1}{\lambda \theta^{\alpha}} \sum_{k=0}^{\infty} \frac{(e^{-\lambda t}(1-\theta) + \theta)^{k+1} - \theta^{k+1}}{(k+1)!} \right. \\
&+ \left. \left(\frac{\theta}{\theta-1} \right)^{\alpha} \frac{1}{\lambda} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \frac{(-1)_l ((e^{-\lambda t}(1-\theta) + \theta)^{l+m-\alpha+1} - \theta^{l+m-\alpha+1})}{l! m! \theta^m (l+m-\alpha+1)} \right).
\end{aligned}$$

Estimation and Inference

Firstly, in order to identify the shape of a lifetime data failure rate function we shall consider a graphical method based on the TTT plot. In its empirical version

the TTT plot is given by

$$G(r/n) = [(\sum_{i=1}^r Y_{i:n}) + (n-r)Y_{r:n}] / (\sum_{i=1}^n Y_{i:n}),$$

where $r=1, \dots, n$ and $Y_{i:n}$ represent the order statistics of the sample. It has been shown that the failure rate function is increasing (decreasing) if the TTT plot is concave (convex).

Assuming the lifetimes are independent to the censoring mechanism, the logarithm of the likelihood function is given by:

$$\begin{aligned}
l(\alpha, \lambda, \theta, y_i) &= \sum_{i=1}^n \delta_i \log \left(\alpha \left(\frac{\theta(1-e^{-\lambda y_i})}{e^{-\lambda y_i}(1-\theta) + \theta} \right)^{\alpha-1} \right. \\
&\times \left. \left(\frac{\theta \lambda e^{-\lambda y_i}}{(e^{-\lambda y_i}(1-\theta) + \theta)^2} \right) \right) - \sum_{i=1}^n \delta_i \log \left(1 - \left(\frac{\theta(1-e^{-\lambda y_i})}{e^{-\lambda y_i}(1-\theta) + \theta} \right)^{\alpha} \right) \\
&+ \sum_{i=1}^n \log \left(1 - \left(\frac{\theta(1-e^{-\lambda y_i})}{e^{-\lambda y_i}(1-\theta) + \theta} \right)^{\alpha} \right). \quad (16)
\end{aligned}$$

The maximum likelihood estimates (MLEs) of the parameters are obtained by direct maximization of the logarithm of the likelihood, where δ_i is an indicator variable, assuming 1 when failure/death occurs and 0 when censure.

We consider the $-l(\cdot)$ values to find the MLE of the parameters using the optim function in R language (R development core Team, 2009), and the Akaike information criterion (AIC) and Bayesian information criterion (BIC), to compare the models, which are defined, respectively, by $-2l(\cdot) + 2q$, and $-2l(\cdot) + q \log(n)$, where $l(\cdot)$ is the Log-Likelihood evaluated in the MLE vector on respective distribution, q is the number of parameters estimated and n is the sample size. The best distribution corresponds to a lower $-l(\cdot)$, AIC and BIC values.

The information matrix is given by:

$$I_F(\alpha, \lambda, \theta) = -E \begin{pmatrix} \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \alpha \partial \alpha} & \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \alpha \partial \lambda} & \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \alpha \partial \theta} \\ \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \lambda \partial \alpha} & \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \lambda \partial \lambda} & \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \lambda \partial \theta} \\ \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \theta \partial \alpha} & \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \theta \partial \lambda} & \frac{\partial^2 l(\alpha, \lambda, \theta)}{\partial \theta \partial \theta} \end{pmatrix}_{(\alpha, \lambda, \theta) = (\hat{\alpha}, \hat{\lambda}, \hat{\theta})}$$

The elements from information matrix can be obtained just numerically.

Data Analysis

In this section our methodology is illustrated in three dataset, two artificial and one real related to mammary tumors in female rats.

Firstly a simulation study was performed to see the behavior of the modeling of different distributions to

the survival times generated from ECEG distribution. We generate two samples of the ECEG distribution by considering the inverse transformation of cumulated density Function (10), Data 1 (D1) and Data 2 (D2), with the same sample size and censoring percentage, $n=20$ and $p=0.1$, respectively. For D1 we consider $(\alpha, \lambda, \theta) = (4.2323, 0.07015, 0.50)$ with increasing hazard, the TTT plot is shown in Figure 3 (left), and for D2 the parametric vector is $(\alpha, \lambda, \theta) = (0.30, 0.0005, 0.975)$ with decreasing hazard (Figure 3, right).

We fit the ECEG distribution to the dataset and compare its fitting with its particular cases, the Complementary Exponential Geometric distribution with pdf given by $f(x) = \lambda \theta e^{-\lambda x} (e^{-\lambda x} (1-\theta) + \theta)^{-2}$, where $\lambda > 0$ is a scale parameter and $0 < \theta < 1$ is a shape parameter and the Exponential (E) distribution with pdf given by $f(x) = 1/\theta e^{-x/\theta}$, where the rate parameter is θ .

The distributions are compared considering the $-l(.) = -\log L(\hat{\alpha}, \hat{\lambda}, \hat{\theta})$ values, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), defined respectively by $-2l(.) + 2q$ and by $-2l(.) + q \log(n)$, where $(\hat{\alpha}, \hat{\lambda}, \hat{\theta})$ are the MLEs vector, q is the number of parameters estimated and n is the sample size. The best distribution corresponds to lower $-l(.)$, AIC and BIC values.

The MLEs of the parameters α , λ and θ of the ECEG distribution are given, respectively, by 4.025, 0.870 and 0.003 for D1, and 0.205, 0.001 and 0.103 for D2.

Table I shows the values of the statistics $-l(.)$, AIC, BIC and the Kolmogorov Smirnov (K-S) statistics with their p-values which are evidence in favor of the model ECEG distribution.

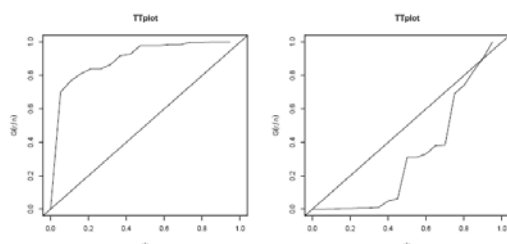


Figure 2 - TTT plot from artificial dataset D1 (left) and D2 (right)

TABLE 1 VALUES OF CRITERION INFORMATION FOR SIMULATED SAMPLE: DATA 1 AND DATA 2

	D1			D2		
	ECEG	CEG	E	ECEG	CEG	E
$-l(.)$	31.129	33.298	57.934	57.934	140.886	138.867
AIC	68.258	70.597	117.869	257.158	285.772	279.735
BIC	71.245	72.589	118.865	260.146	287.764	280.731
K-S	0.197	0.253	0.423	0.214	0.429	0.410
p-value	0.414	0.151	0.001	0.317	0.001	0.002

The Figure 4 shows the Survival curve of the fitted models, corroborating the values in Table 1.

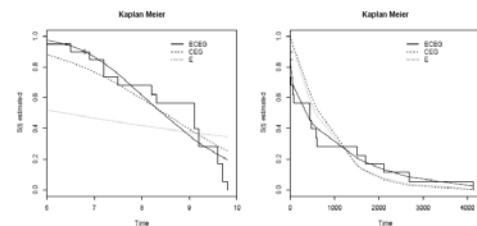


Figure 3 - Fitted survival functions of the ECEG, CEG, E distributions superimposed to the Kaplan-Meier fit from Data 1 (left) and Data 2 (right)

As a real example consider the dataset is from Lawless (2003), which consists in data from a nine-month study on the effect of known carcinogens DES and DMBA in the induction of mammary tumors in female rats (Shellabarger, McKnight, Stone and Holtzman, 1980).

Table 2 shows the parameter MLEs according to each one of the five fitted distributions, and the values of the AIC, BIC and $-l(.)$, and also the K-S statistics with their p-values. The ECEG distribution outperforms its concurrent distributions in all considered criterion, corroborating the fact that the ECEG distribution can be seen as a competitive distribution of practical significance for the analysis of survival data.

These conclusions are corroborated by the fitted survival functions of the ECEG, CEG, E distributions superimposed to the Kaplan-Meier fit. We observed a clear difference between the fitted curves, which is a strong motivation for choosing the most suitable distribution for fitting the data.

TABLE 2 ESTIMATIVE AND VALUES OF CRITERION INFORMATION FROM REAL DATASET

Mode l	Parameter s (α, λ, θ)	AIC	BIC	$-l(.)$	K-S Stat	P-value
ECEG	21.091; 0.023; 0.971	211.358	214.765	102.679	0.214	0.238
CEG	-; 0.030; 0.011	215.959	218.230	105.979	0.225	0.194
E	-; 175.860 ; -	236.455	237.591	117.227	0.263	0.082

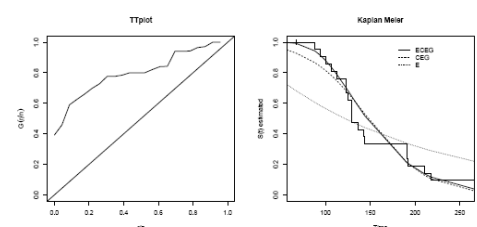


Figure 4 - Real data: TTT plot (left) and fitted survival functions of the ECEG, CEG, E distributions superimposed to the Kaplan-Meier fit (right)

Conclusion Remarks

In this paper we introduce an extension of the

Complementary Exponential Geometric distribution. The new distribution is much more flexible than its predecessor CEG distribution, presenting increasing and decreasing shaped failure rates. We provide statistical properties of the ECEG distribution including reliability measures, the density, failure rate, moments, quantiles, order statistics. Estimation via maximum likelihood is straightforward. Artificial and real data applications of the ECEG distribution shows that it could provide a better fit than its particular cases.

REFERENCES

- Adamidis, K. and Loukas, S. "A lifetime distribution with decreasing failure rate". *Statistics Probability Letters*, 39: 35–42, 1998.
- Bakouch, H. S., Al-Zahrani, B. M., Al-Shomrani, A. A., Marchi, V. A. A. and Louzada, F. "An extended lindley distribution". *Journal of the Korean Statistical Society*, 2011.
- Barreto-Souza, W. and Cribari-Neto, F. "A generalization of the Exponential-Poisson distribution". *Statistics and Probability Letters*, 79: 2493–2500, 2009.
- Barreto-Souza, W., Morais, A. L. and Cordeiro, G. M. "The Weibull-Geometric distribution". *Journal of Statistics Computation and Simulation*, 60:35–42, 2010.
- Barriga, G. D. C., Louzada, F. and Cancho, V. G. "The Complementary Exponential Power lifetime model". *Computational Statistics and Data Analysis*, 54(5): 1250–1259, 2011.
- Basu, A. and Klein, J. "Some recent development in competing risks theory". In *Survival Analysis*, Edited by Crowley, J. and Johnson, R.A., Hayward: IMS, 1(1), 216–229, 1982.
- Cancho, V. G., Louzada, F. and Barriga, G. D. C. "The Poisson-Exponential lifetime distribution". *Computational Statistics & Data Analysis*, 55(1): 677–686, 2011.
- Cox, D.R. and Oakes, D. "Analysis of survival data". Chapman and Hall, London, 1984.
- Crowder, M. J., Kimber, A. C., Smith, R.L. and Sweeting, T.J. "Statistical analysis of reliability data". Chapman and Hall, London, 1991.
- Glaser, R. E. "Bathtub and related failure rate characterizations". *Journal of the American Statistical Association*, 75: 667–672, 1980.
- Gupta, R. and Kundu, D. "Generalized Exponential distributions". *Australian and New Zealand Journal of Statistics*, 41: 173–188, 1999.
- Lawless, J. F. "Statistical models and methods for life-time data". Wiley, New York, NY, 2nd edition, 2003.
- Louzada, F., Roman, M. and Cancho, V. G. "The Complementary Exponential Geometric distribution: Model, properties, and a comparison with its counterpart". *Computational Statistics & Data Analysis*, 55(8): 2516–2524, 2011.
- Marshall, A. W. and Olkin, I. "Life distributions". Springer, 2007.
- Mudholkar, G. S., Srivastava, D. K., and Freimer, M. "The Exponentiated Weibull family: A reanalysis of the bus-motor-failure data". *Technometrics*, 37(4): 436–445, 1995.
- Nadarajah, S. and Kotz, S. "The exponentiated type distributions". *Acta Applicandae Mathematicae*, 92(2): 97–111, 2006.
- Shellabarger, C. J., McKnight, B., Stone, J.P., and Holtzman, S. "Interaction of dimethylbenzanthracene and diethylstilbestrol on mammary adenocarcinoma formation in female ACI rats". *Cancer Res.*, 40, 1808–181, 1980.
- Silva, G. O., Ortega, E. M. M and Cordeiro, G. M. "The Beta Modified Weibull distribution". *Lifetime Data analysis*, Doi 10.1007/s10985-010-9161-1, 2010.
- Silva, R. B., Barreto-Souza, W. and Cordeiro, G. M. "A new distribution with decreasing, increasing and up-side-down bathtub failure rate". *Computational Statistics & Data analysis*, 69: 109–124, 2010.
- Smith, R. M. and Bain, L. J. "An Exponential Power life-testing distribution". *Communications in Statistics, Theory and Methods*, 4: 469–481, 1975.

Appendix

Through the paper we shall use the following equations:

$$\int_0^1 z^{p-1} (1-z)^{n-1} (1+bz^m)^l dz = \Gamma(n) \sum_{k=0}^{\infty} \binom{l}{k} \frac{b^k \Gamma(p+km)}{\Gamma(p+n+km)}, \quad (17)$$

where $\Gamma(a,b)$ is the incomplete Gamma function given by $\Gamma(a,b) = \int_b^{\infty} t^{a-1} e^{-t} dt$.

$$(1-x)^{-r} = \sum_{k=0}^{\infty} \frac{(r)_k}{k!} x^k, \quad (18)$$

where $(r)_k$ is a Pochhammer symbol, given $(r)_k = r(r+1) \dots (r-k+1)$ and if $|x| < 1$ the series converge, and $(-r)_k = (-1)^k (r-k+1)_k$.